American Journal of Medical Genetics 64:216–219 (1996)

Extended Gene Diversity at the FMR1 Locus and Neighbouring CA Repeats in a Sub-Saharan Population

P. Chiurazzi, G. Destro-Bisol, M. Genuardi, B.A. Oostra, G. Spedini, and G. Neri

Istituto di Genetica Medica, Università Cattolica and Centro Ricerche per la Disabilità Mentale e Motoria, Associazione Anni Verdi, Roma (P.C., M.G., G.N.); Dipartimento di Biologia Animale e dell' Uomo, Università La Sapienza, Roma (G.D.-B., G.S.), Italy; Department of Clinical Genetics, Erasmus University, Rotterdam, The Netherlands (P.C., B.A.O.)

We report on the allele distributions in a normal black African population at two microsatellite loci neighbouring the FRAXA locus and at the CGG repeat in the 5′ end of the FMR-1 gene, which causes the fragile X syndrome. The CGG repeat distribution was found to be similar to that of other ethnic groups, as well as to that of other non-human primates, possibly predicting a comparable prevalence of fragile X in Africa. Significant linkage disequilibrium has been observed between fragile X mutations and alleles of the DXS548 and FRAXAC1 loci in European and Asian populations, and some founder chromosomes may be extremely old. Those associated with FRAXAC1-A and DXS548-2 alleles are not present in the Asian fragile X samples. We searched for these alleles and their frequency in the well defined Bamileke population of Cameroon. All previously described alleles and some new ones were found in this sample, supporting the hypothesis of their pre-existence and subsequent loss in Asian populations. Finally, the heterozygosity of the Bamileke sample was significantly higher at both marker loci and comparable to that of Europeans at the CGG repeat, confirming the notion that genetic diversity is greater in Africans than in other groups and supporting the view that evolution of modern man started in Africa.

KEY WORDS: gene diversity, FMR1 locus microsatellite markers, Africa

INTRODUCTION

Few reports have dealt with fragile X in sub-Saharan Africa [Sutherland and Hecht, 1985; Venter et al., 1981; Howard-Peebles and Stoddard, 1980]; and, although the prevalence estimate in whites of European origin has been recently lowered to approximately 1 in 4,000 males [Turner et al., 1996], fragile X syndrome seems either to be less frequent or to have been underascertained in African populations. Watkins et al. [1995] surveyed five trinucleotide repeat loci other than fragile X and suggested that the presence of alleles at the upper end of the normal range might correlate with an increased prevalence of the corresponding diseases, as in the case of dentato-rubro-pallido-luysian atrophy (DRPLA) in the Japanese population. Thus, we determined the allele distribution of the CGG repeat at the 5′ end of the FMR1 gene in a normal sample from the Bamileke population of Cameroon, in order to test whether a possible lack of larger alleles might explain a low incidence of fragile X in Africa, as was proposed by Goldman et al. [1994] in the case of myotonic dystrophy. We also tested this population with flanking markers FRAXAC1 and DXS548, to verify whether alleles FRAXAC1-A and DXS548-2, which are frequently found on European FMR1 founder chromosomes [Chiurazzi et al., 1996; Macpherson et al., 1994], are present in Africa. These alleles are completely missing in the Japanese and Chinese patients populations [Richards et al., 1994; Arinami et al., 1993; Zhong et al., 1994] and we hypothesize their preexistence in Africa, their subsequent spread to Indoeuropeans and loss in the Asian populations.

SAMPLE AND METHODS

Our sub-Saharan African sample is comprised of 85 independent chromosomes derived from 9 male and 38 female unrelated Bamileke individuals. Bamilekes are considered the major ethnic group of Cameroon, accounting for 11% of the country population [Spedini and Destro-Bisol, 1988]. An additional 8 male Bororo and 8

Received for publication October 20, 1995; revision received December 20, 1995.

Address reprint requests to Giovanni Neri, MD, Istituto di Genetica Medica, Università Cattolica, L.go F. Vito 1, 00168 Rome, Italy.

male Sanga individuals from the Central African Republic were also tested. Bororos, considered the most unadmixed group of the larger Fulbe population, are nomadic shepherds, while the Sanga people derive their name from the river on which they base their subsistence. According to Greenberg [1980], the languages spoken by these populations belong to the same linguistic philum (Niger-Kordofan), although the linguistic family and group shared by Bamilekes and Sangas (Bantoid-Bantu) differ from those of Bororos (West Atlantic-Fulfulde). The microsatellites studied were FRAXAC1 [Richards et al., 1991] and DXS548 [Riggins et al., 1992] which are located 7 kbp and 150 kbp, respectively, proximal to the CGG repeat at the 5′ end of the FMR1 gene. A duplex PCR protocol has been developed to allow simultaneous amplification in a single reaction [Chiurazzi et al., 1996]. Primers "c" and "f" [Fu et al., 1991] were employed to assess the CGG repeat number with 200 ng genomic DNA, 0.65 units of Taq polymerase, 2.0 mM $MgCl_2$, 10% DMSO, 0.2 mM each dNTP—except dGTP which was 0.05 mM—0.15 mM 7-deaza-dGTP, 0.13 μl of alpha-^{32}P-dCTP (3,000 Ci/ml) and 2 pmol of each primer. Thirty cycles were again used in a two-step PCR (94°/45 sec-68°/2 min 30 sec). Samples were run at 70 W for 2–4 hr, depending on the size of the amplified products (approximately 200 and 150 bp for the duplex PCR and 300 bp for the 30 CGG repeat allele). Gels were dried without fixation and exposed for 12 to 48 hr at −80°C with intensifying screens. Reference DNAs were consistently re-loaded at the sides and in the middle of every gel to rule out gel distortion artifacts and samples were reamplified if they failed to give any signal the first time or if they showed ambiguous bands. CGG repeat size was evaluated next to an M13 sequence assuming the thickest band to be the original product, as 2–3 faster shadow bands were always present. FRAXAC1 alleles were named with letters (A–F) as in Richards et al. [1992] and DXS548 alleles were named as in MacPherson et al. [1994]. Heterozygosity and its variance were estimated for all three ethnic groups using formulas [8.4] and [8.12] described in Nei [1987].

RESULTS

Because of the low number of Sanga and Bororo individuals tested, only data from the Bamileke sample have been thoroughly analyzed and are presented in Figure 1 and Table I. Italian [Chiurazzi et al., 1996] and Chinese [Zhong et al., 1994] control groups are also displayed to provide comparison with Caucasian and Asian populations, respectively. Figure 1a represents the allele distribution at the CGG repeat of the FMR1 gene and clearly shows that the African distribution has two modes, with a substantial proportion of alleles at 30 repeats (27%) and at 29 repeats (27%), coinciding with the European and Asian modes, respectively. A minor peak is present at 22 repeats (7%) and the range spans from allele 22 to 41 (Italian range, 18–47; Chinese range, 21–45), while 15.3% of the sample is above 31 repeats (24.1% in Italians; 19.5% in Chinese).

Figure 1b and c show the distributions of FRAXAC1 and DXS548 alleles, which are rather spread out in the African population. At the FRAXAC1 locus, again the Bamileke chromosomes are almost equally distributed between the European mode (allele C, 53.6%) and the Asian one (allele D, 37.8%) and the range comprises alleles A to F, this latter being described before only in one fragile X patient from Southern Italy [Chiurazzi et al., 1996]. At the DXS548 locus one can notice rare alleles too, as DXS548-3 (11%) which we found in almost 9% of our fragile X patients but not in controls [Chiurazzi et al., 1996], and DXS548-0, that we observed in a Sanga male, which has never been reported before. As we previously noted [Chiurazzi et al., 1996], intermediate-size alleles 1.5, 3.5 and 5.5 (1, 4 and 3 cases, respectively) could be due to a 1 bp insertion/deletion polymorphism within a variable $(C)_4G(C)_{11}$ sequence, approximately 50 bp upstream from the GT repeat.

Table I compares the expected heterozygosities at every locus, demonstrating that the Bamilekes have the highest gene diversity at both marker loci FRAXAC1 and DXS548, while the estimated value intervals (± standard error) of the Bamilekes and Italians extensively overlap at the FMR1 CGG repeat locus. Finally, we will also note that DXS548 intermediate-size alleles are more frequent in Bamilekes (9.7%) than in Italians, where they can be considered extremely rare (1 out of 215, i.e., <0.5%); thus we may assume that Africans show an increased heterozygosity also at the $(C)_4G(C)_{11}$ contained in the DXS548 PCR product. Therefore, we are currently investigating the variability of this other polymorphic element by using a different 5′ primer. This will help to exploit the full informative potential of the main GT repeat without blurring effects.

DISCUSSION

Surveys of allele distribution at various microsatellite loci in different human populations and groups of non-human primates have helped to elucidate their mutational mechanism, while at the same time shedding some light on the evolutionary history of our species [Cavalli-Sforza, 1991; Cavalli-Sforza and Piazza, 1993]. A complex stepwise mutational mechanism (SMM), allowing both expansion and reduction to occur mainly by single bp slippage, but also in discrete "jumps," seems to fit best the experimental data [Di Rienzo et al., 1994; Deka et al., 1995] accounting for the multimodal distributions often observed at several loci [Mandel, 1994]. If constant mutation rates and no selective constraints are assumed for microsatellite variation, gene diversity can be correlated with time elapsed since foundation, although a small effective population size (N_e) could reduce the observed variability [Paabo, 1995]. We observed the highest heterozygosity in the ethnically well defined Bamilekes (Table I), despite the larger size of the two non-African samples and their presumably greater extent of recent admixture. This observation is consistent with other evidence based on PCR analysis of mini- and microsatellite loci in various human populations [Bowcock et al., 1994; Destro-Bisol et al., 1994; Deka et al., 1995]. Although the evolutionary significance of a greater genetic diversity is controversial [Rogers and Jorde, 1995], our data might support the hypothesis of

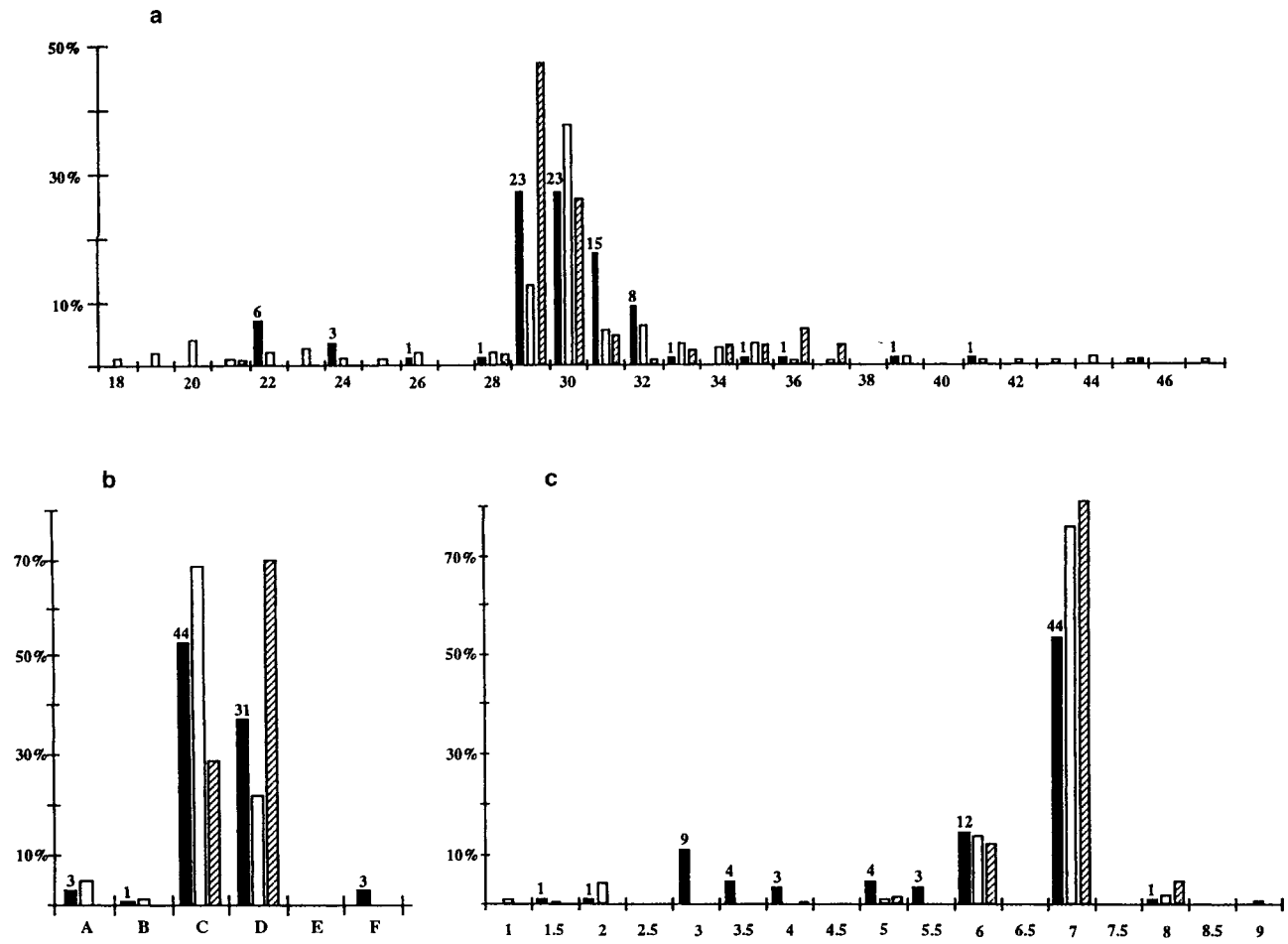

Fig. 1. **a:** CGG repeat number distribution at FMR1 locus of 85 unrelated normal chromosomes from the Bamileke ethnic group of Cameroon (black bars). Allele distribution at the **(b)** FRAXAC1 locus and **(c)** DXS548 locus of 82 unrelated normal chromosomes from the same population. Absolute numbers are reported next to every bar; and, for comparison, European [Chiurazzi et al., 1996] and Asian [Zhong et al., 1994] control populations are represented by white and shaded bars, respectively.

a monocentric origin of modern man, migrating out of Africa [Gibbons, 1995].

The substantially similar distribution of Bamileke, Italian, and Chinese alleles at the FMR1 CGG repeat locus (Fig. 1a), with the highest observed African allele at 41 repeats (vs. 47 in Italians and 45 in Chinese), would suggest a detectable prevalence of the fragile X syndrome in Africa. On the other hand, Rubinsztein et al. [1995] have proposed the existence of a general trend toward size increase of microsatellite repeats during evolution, based on comparison between human and primate samples, and indeed the lack of high-normal range alleles at the CAG repeat of the myotonin locus might explain why the prevalence of myotonic dystrophy in Africa appears low [Goldman et al., 1994; Watkins et al., 1995]. Deelen et al. [1994] first reported CGG counting at the FMR1 locus in an African green monkey that showed two alleles of 26 and 31 repeats, while Zhong et al. [1995] tested 8 chimpanzees and 6 orangutans, finding a mean repeat length of 34 and 30 repeats, respectively, and also confirming the presence of intercalated AGGs. Actually, the same data of

TABLE I. Expected Heterozygosity Values*

	Bamileke	Italians	Chinese
FMR1-CGG	0.816 ± 0.021 (85)	0.832 ± 0.027 (141)	0.706 ± 0.033 (123)
FRAXAC1	0.573 ± 0.032 (82)	0.467 ± 0.031 (231)	0.419 ± 0.026 (206)
DXS548	0.679 ± 0.050 (82)	0.405 ± 0.039 (211)	0.327 ± 0.039 (206)

* Heterozygosity values and standard errors at the three analyzed loci were calculated for the Bamileke ethnic group of Cameroon [present study], the Italian control sample that we recently reported [Chiurazzi et al., 1996] and the Chinese studied by Zhong et al. [1994]. Total chromosome numbers are indicated in brackets.

Rubinsztein et al. [1995] demonstrate a complete overlap of human FMR1 CGG allele distribution with that of gorillas, orangutans and baboons, while their sample of 19 chimpanzees showed an even higher mean length. So it seems more likely to assume that microsatellite sizes fluctuate during evolution in balance between expansion and reduction trends, and the FMR1 CGG repeat itself seems to have reached its present size already before the other primates diverged from the human lineage. As further proof, we should consider that our African sample included two previously undescribed alleles, one (FRAXAC1-F) being the smallest at its locus and the other (DXS548-0) being the largest; gene diversity in Africans is then increased both in the sense of increasing and decreasing size. From Figure 1b and c we verify the presence in the Bamilekes of both alleles FRAXAC1-A and DXS548-2 which are frequently associated on fragile X chromosomes in Caucasians, while they are almost absent in the Far East Asians. They were also found in two Iranian and one Singhalese patients that were referred to us [Chiurazzi et al., 1996]. Thus we speculate that Indoeuropean fragile X A-2 chromosomes could derive from an ancient African founder, while they might have been lost in Asian populations, possibly because of a "bottleneck" effect.

ACKNOWLEDGMENTS

This work was supported in part by contribution from MURST and also from C.N.R. funds to G.S. (research project: "Basi tassonomiche, corologiche ed evolutive per lo studio della diversità animale"). P.C. wishes to express his gratitude to Dr. Esther de Graaff and Prof. Hans Galjaard for their generous support and gratefully acknowledges the financial support of a C.N.R. grant for his travel and stay in Rotterdam. We also especially thank Prof. Alberto Piazza and Prof. Francesco Andreasi-Bassi for their helpful comments.

REFERENCES

Arinami T, Asano M, Kobayashi K, Yanagi H, Hamaguchi H (1993): Data on the CGG repeat at the fragile X site in the non-retarded Japanese population and family suggest the presence of a subgroup of normal alleles predisposing to mutate. Hum Genet 92: 431–436.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994): High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457.

Cavalli-Sforza LL (1991): Genes, peoples and languages. Sci Am 265:72–78.

Cavalli-Sforza LL, Piazza A (1993): Human genomic diversity in Europe: A summary of recent research and prospects for the future. Eur J Hum Genet 1:3–18.

Chiurazzi P, Genuardi M, Kozak L, Giovannucci-Uzielli ML, Bussani C, Dagna-Bricarelli F, Grasso M, Perrone L, Sebastio G, Sperandeo MP, Oostra BA, Neri G (1996): Fragile X founder chromosomes in Italy: Few initial events and possible explanation for their heterogeneity. Am J Med Genet 64:209–215.

Deelen W, Bakker C, Halley DJJ, Oostra BA (1994): Conservation of CGG region in FMR1 gene in mammals. Am J Med Genet 51:513–516.

Deka R, Jin L, Shriver MD, Yu LM, DeCroo S, Hundrieser J, Bunker CH, Ferrell RE, Chakraborty R (1995): Population genetics of dinucleotide (CA)(GT) polymorphism in world populations. Am J Hum Genet 56:461–474.

Destro-Bisol G, Presciuttini S, d'Aloja E, Dobosz M, Spedini G, Pascali VL (1994): Genetic variation at the ApoB 3' HVR, D2S44, and D7S21 loci in the Ewondo ethnic group of Cameroon. Am J Hum Genet 55:168–174.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994): Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci 91:3166–3170.

Fu YH, Kuhl DPA, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJMH, Holden JJA, Fenwick RG, Warren ST, Oostra BA, Nelson DL, Caskey CT (1991): Variation of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman paradox. Cell 67:1047–1058.

Gibbons A (1995): Out of Africa—at last? Science 267:1272–1273.

Goldman A, Ramsay M, Jenkins T (1994): Absence of myotonic dystrophy in southern African negroids is associated with a significantly lower number of CTG trinucleotide repeats. J Med Genet 31:37–40.

Greenberg JH (1980): Classification des langues d'Afrique. In J. Ki Zerbo (ed): "Histoire generale d'Afrique" tome I. Paris: Jeune Afrique/Unesco, pp 321–338.

Howard-Peebles PN, Stoddard GR (1980): Race distribution in X-linked mental retardation with macro-orchidism and fragile site in Xq. Am J Hum Genet 32:629–630.

Macpherson JN, Bullman H, Youings SA, Jacobs PA (1994): Insert size and flanking haplotype in fragile X and normal populations: possible multiple origins for the fragile X mutation. Hum Mol Genet 3:399–405.

Mandel JL (1994): Trinucleotide diseases on the rise. Nat Genet 7: 453–455.

Nei M (1987): "Molecular Evolutionary Genetics." New York: Columbia University Press.

Paabo S (1995): The Y chromosome and the origin of all of us (men). Science 268:1141–1142.

Richards RI, Holman K, Kozman H, Kremer E, Lynch M, Pritchard M, Yu S, Mulley J, Sutherland GR (1991): Fragile X syndrome: genetic localization by linkage mapping of two microsatellite repeats FRAXAC1 and FRAXAC2 which immediately flank the fragile site. J Med Genet 28:818–823.

Richards RI, Holman K, Friend K, Kremer E, Hillen D, Staples A, Brown WT, Goonewardena P, Tarleton J, Schwartz C, Sutherland GR (1992): Evidence of founder chromosomes in fragile X syndrome. Nat Genet 1:257–260.

Richards RI, Kondo I, Holman Kn K, Yamauchi M, Seki N, Kishi K, Staples A, Sutherland GR, Hori T (1994): Haplotype analysis at the FRAXA locus in the Japanese population. Am J Med Genet 51:412–416.

Riggins GJ, Sherman SL, Oostra BA, Sutcliffe JS, Feitell D, Nelson DL, van Oost BA, Smits APT, Ramos FJ, Pfender E, Kuhl DPA, Caskey CT, Warren ST (1992): Characterization of a highly polymorphic dinucleotide repeat 150 kb proximal to the fragile X site. Am J Med Genet 43:237–243.

Rogers AR, Jorde LB (1995): Genetic evidence of modern human origins. Human Biology 67:1–36.

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li SH, Margolis RL, Ross CA, Ferguson-Smith MA (1995): Microsatellite evolution—evidence for directionality and variation in rate between species. Nat Genet 10:337–343.

Spedini G, Destro-Bisol G (1988): Etude anthropologique au Cameroun. Le peuplement ancien. Données archeologiques, linguistiques at anthropologiques. Antropologia Contemporanea 11:93–112.

Sutherland GR, Hecht F (1985): Racial distribution. In: "Fragile sites on human chromosomes." Oxford University Press, p 177.

Turner G, Webb T, Wake S, Robinson H (1996): The prevalence of the fragile X syndrome. Am J Med Genet 64:196–197.

Venter PA, Gericke GS, Dawson B, Op't Hof J (1981): A marker X chromosome associated with nonspecific male mental retardation. S Afr Med J 60:807–811.

Watkins WS, Bamshad M, Jorde LB (1995): Population genetics of trinucleotide repeat polymorphism. Hum Mol Genet 4:1485–1491.

Zhong N, Liu X, Gou S, Houck GE Jr, Li S, Dobkin C, Brown WT (1994): Distribution of FMR1 and associated microsatellite alleles in a normal Chinese population. Am J Med Genet 51:417–422.

Zhong N, Yang W, Dobkin C, Brown WT (1995): Fragile X gene instability: Anchoring AGGs and linked microsatellites. Am J Hum Genet 57:351–361.